Multiple Analysis of Remarks of Elderly and Disabled People by Text Mining

Yamada, Yasuhiro
Research Institute for Information Technology, Kyushu University

Kato, Kanji
User Science Institute, Kyushu University

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

出版情報：International Conference on Kansei Engineering and Emotion Research 2007(KEER2007), 2007
バージョン：

Multiple Analysis of Remarks of Elderly and Disabled People by Text Mining

Yasuhiro YAMADA*, Kanji KATOH** and Sachio HIROKAWA*

*Research Institute for Information Technology, Kyushu University, Fukuoka-shi Fukuoka 812-8581, JAPAN, {yamada,hirokawa}@cc.kyushu-u.ac.jp

** Kyushu University User Science Institute, Fukuoka-shi Fukuoka 812-8581, JAPAN, kanji@usi.kyushu-u.ac.jp

Abstract: This paper describes a method to analyze needs from the remarks of elderly and disabled people by using two text mining systems, which are the concept graph system and matrix search system. First, we collected 2,409 remarks about products and services from 19 people. Next, we transformed remarks with meaningful content into an evaluation record with ten aspects, and made 680 evaluation records. The two text mining systems analyze evaluation records in which an input keyword is used in a manner similar to search engines. The concept graph system visualizes a hierarchy of feature words in the records. We can understand important words and their relations in the records before we read the records. The matrix search system displays a matrix of the distribution of clustering from two aspects of the records. We can understand the relation of two aspects in the records by combining various aspects. The two systems achieve multiple and dialogical analysis of evaluation records. Therefore, the systems enable us to examine various problems that we choose to target. In experiments, we made an analysis of the records by combining the results of the two text mining systems.

***Keywords*:** *Concept Graph, Matrix Search, Needs Analysis, Text Mining*

1. Introduction

The environment of modern society is not easy to live in for many elderly and disabled people. They want to improve the inconvenience of their living environment and want to live without restrictions. Conversely, businesses and government want to improve their environment and develop products for them. Government policies concerning elderly people are especially important because Japan is becoming a rapidly aging society. It is necessary to understand their various needs. Examples of methods to extract the needs are interviews and questionnaires.

In order to study and develop improved living conditions for elderly and disabled people, we can analyze data collected from interviews and questionnaires. However, the needs are varied and dependent on individual kansei. In order to discern the various needs which are common to all of them or which are dependent on an individual, a method is needed to analyze the data from various viewpoints.

In this study, we consider the situation in which analysts examine various problems using the data: for example, analyses of the needs of a person, the needs of people with a disability, and their needs for a product. With a specialized system which is given a keyword as input and outputs the result like search engines, we can analyze the data dialogically.

The evaluation grid method is an interview research method which visualizes user needs structurally [1]. On the other hand, there has been research which summarizes collected interview remarks and analyzes the summarized data by text mining. The graph theorem, Bayesian network and KeyGraph [2] are used to analyze the data.

DIAMiningEX [3] is a system that analyzes texts such as questionnaires. It retrieves texts and analyzes them dialogically. The system has a word co-occurrence table, which is a list of word frequencies. There exist many text mining systems in addition to DIAMiningEX.

This paper proposes a method to analyze the needs from remarks collected from interviews of elderly and disabled people by using a concept graph system [4,5] and a matrix search system [6-8].

In order to analyze the needs based on user language, we first collected 2,409 remarks from 19 people, who included elderly, disabled or able-bodied people. We

Table 1. Elderly and disabled peoples' remarks

It is difficult to recognize a switch because it stays at the same position when I turn the power on or off.
It is helpful to add a switch to the same beeps of different MD players when I play, stop, rewind, fast-forward, and record music on them.
I prefer a switch with an announce function to it with a beep.
I can recognize a switch easily if it indents when I push it. But, there exist recently a lot of types of switches which stay at the same position when I turn the power on or off. It is difficult for me to recognize this type of switch.

asked them the reasons why they like or dislike some products and services. We also asked about their living activities and opinions when they use them. In order to achieve a multiple analysis of the remarks, we transformed remarks with meaningful content into an evaluation record with ten aspects, and made 680 evaluation records from all the remarks.

Next, we analyzed the records by using a concept graph system and a matrix search system. A common search engine, such as Google or Yahoo!, outputs a list of texts in which an input keyword appears. The two text mining systems analyze the evaluation records of a search result. Therefore, the systems achieve a dialogical analysis of the records by inputting various keywords that we want to examine.

A concept graph system visualizes a hierarchy of feature words in the records. We can understand important words and their relations in the records before we read the records. A matrix search system displays a matrix of the distribution of the clustering from two aspects of the records. We can know the relation of two aspects in the results. The various combinations of two aspects achieve a multiple analysis of the records. The matrix search system can handle non-numeric data such as sentences, compared to OLAP [9], which handles numeric data only.

We briefly described a method to analyze needs using the two text mining systems in [10]. In this paper, we increase the number of evaluation records and improve the systems. We show experiments of extracting needs from the records and evaluate the systems qualitatively.

In the experiments, we examine (1) the common needs, whose target is a product and a service, among multiple people and (2) individual needs. We make an analysis of the records by combining the results of the two systems.

The collected remarks and evaluation records are written in Japanese. Note that all figures and tables in this paper were originally written in Japanese. We translated the sentences and words in the figure and tables into English. However, some words which are difficult to translate are deleted in the figures and tables.

This paper is organized as follows. Section 2 describes the evaluation records which were generated from the elderly and disabled peoples' remarks. Section 3 introduces the two text mining systems, which are a concept graph system and a matrix search system, to analyze the records. Section 4 describes experiments of the analysis of the records. Section 5 concludes the paper.

2. Elderly and Disabled Peoples' Remarks

This section describes the remarks of elderly and disabled people and evaluation records which were transformed from the remarks.

First, we collected 2,409 remarks from 19 people, who included elderly, disabled or able-bodied people. The disabilities of the disabled people included completely blind, rheumatism, deafness and blindness, deaf-mute, and so on. We asked them about 147 products and services, such as a mobile phone, a remote control and a train station. We asked them the reason why they like or dislike these products and services. We also asked about their living activities and opinions when they use these products and services. Table 1 shows some of the remarks about a switch.

Next, in order to analyze the remarks multilaterally, we transformed remarks with meaningful content into an evaluation record with the following ten aspects, and made 680 evaluation records from all the remarks. If a remark had multiple contents, multiple records were generated from the remark. The records were generated manually by extracting keywords and sentences related to the ten aspects. Table 2 shows some of the evaluation records for a switch.

(1) person
    an interviewee
(2) disability
    a kind of disability
(3) operation / property
    restriction by disability

Table 2. Elderly and disabled peoples' evaluation records

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
completely blind woman	completely blind	checking by touch and hearing				remote control etc.	switch	I prefer a switch with an announce function.	I don't prefer a switch with only beep because I don't know how to use it.
completely blind woman	completely blind				remote control operation	remote control etc.	switch	I prefer a switch with sign-on.	
completely blind woman	completely blind	checking by touch and hearing				remote control etc.	switch	I prefer a switch with the sign of sound.	I don't prefer a switch which stays at the same position when I turn the power on or off.
completely blind man	completely blind	checking by senses except for sight		to use with safety		household electrical goods	switch	I prefer a switch with sound.	I don't prefer a switch with one-touch operation because it is too easy to use it.
completely blind man	completely blind	checking by touching				household electrical goods	switch	I prefer a switch whose button indents because it is easy to recognize it.	I don't prefer a switch which stays in the same state when I turn the power on or off.

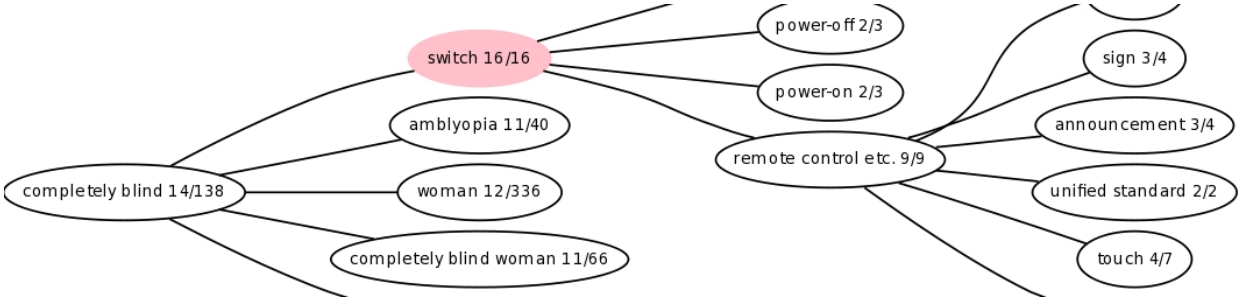

Fig. 1 Part of the concept graph of evaluation records in which the word "switch" appears

(4) evaluation factor 1
a factor which is common to interviewee's life
(5) evaluation factor 2
a factor which is common to other products and services
(6) evaluation factor 3
a factor in only the product and service
(7) product / service
a target product and service for evaluation
(8) part / function
a part and function of a product and service
(9) positive keyword
a keyword based on a positive evaluation
(10) negative keyword
a keyword based on a negative evaluation

We could not extract keywords and sentences related to the ten aspects from some remarks. The number of extracted data in (1) is 680 out of 680 records, (2) is 312, (3) is 219, (4) is 220, (5) is 177, (6) is 167, (7) is 498, (8) is 218, (9) is 434, and (10) is 269.

3. Text Mining Systems for Analysis

This section introduces the concept graph system and matrix search system which is used to analyze the evaluation records of elderly and disabled people. Both systems are given a search keyword for analyzing documents in which the keyword appears. Both systems use Generic Engine for Transposable Association (GETA)[1] for indexing and searching the documents.

3.1 Concept Graph System

A concept graph system [4,5] visualizes the hierarchy of feature words which appear in the retrieval documents. The hierarchy is generated dynamically. We can recognize the key points of words in the documents from a bird's-eye view without reading the documents.

We describe now the definition of a feature word and the hierarchy. Let D be a set of documents, and w be a word. The number of documents in which w appears is denoted by df(w, D). Let u and v be words. The number of documents in which both u and v appear is denoted by df(u*v, D). Let U be a set of all documents and D be a

[1] http://geta.ex.nii.ac.jp

subset of U. A word w is a feature word of D in the condition that df(w, D)/df(w, U) > x. A word w is superior to a word v in the condition that df(u*v, D)/df(v, D) > y and df(u, D) > df(v, D). The parameters x and y are set on the interface of the system.

The concept graph system has been implemented for activities of academic staff [5], music play lists [4], paper lists of conferences, and an English-Japanese dictionary.

Fig. 1 shows part of the concept graph of the evaluation records in which the word “switch” appears. We see an evaluation record as a document. We set the parameters x and y to 0.5 and 0.5. The denominator in a node stands for the document frequency of the word in the node in all records. The numerator stands for the document frequency of the word in the retrieval records. The number of retrieval records is 16. The left word is superior to the right words. For example, we see that the word “completely blind” is superior to “switch.” This indicates that most people talking about a switch have completely blind. Actually, 88% of the evaluation records of a switch describe completely blind. We also see that “remote control etc.” is superior to “sign,” “announcement,” “unified standard” and “touch.” This indicates that most people talking about these four words talk about remote control etc. in the records for a switch. The features of Fig. 1 include the features of Table 2.

3.2 Matrix Search System

The matrix search system [6-8] clusters retrieval documents from two aspects selected by the user. The result is displayed as a two-dimensional table in which documents with the same (similar) contents are assigned

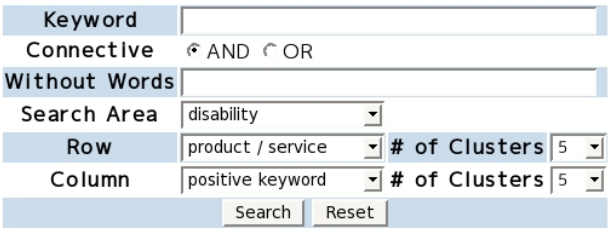

Fig. 2 The interface of the matrix search engine

to a cell. We can know the relation of two aspects by combining various aspects.

Fig. 2 shows the interface of the system. A user inputs a search keyword into the textbox “Keyword” and selects the target search aspect in the pull-down menu of “Search Area.” The user also selects two aspects for clustering and the number of clusters in the pull-down menus of “Row” and “Column.” Selecting a target search aspect and two aspects for clustering achieves a multiple analysis of documents.

Fig. 3 shows the output of the system. We see an evaluation record as a document. The number of records assigned to a cell is displayed in the table. We can understand the amount of need by the number. The feature words of each cluster of each axis are also displayed. We can see the features of a cluster by the words without reading the records.

In Fig. 3, the search keyword is “switch” and we selected “part / function” as the target search aspect, “product / service” as the vertical axis, and “negative keyword” as the horizontal axis. We see that the same needs exist. These needs are about “power-on,” “power-off” and “position” between “remote control etc.” and “household electrical goods.” The number after

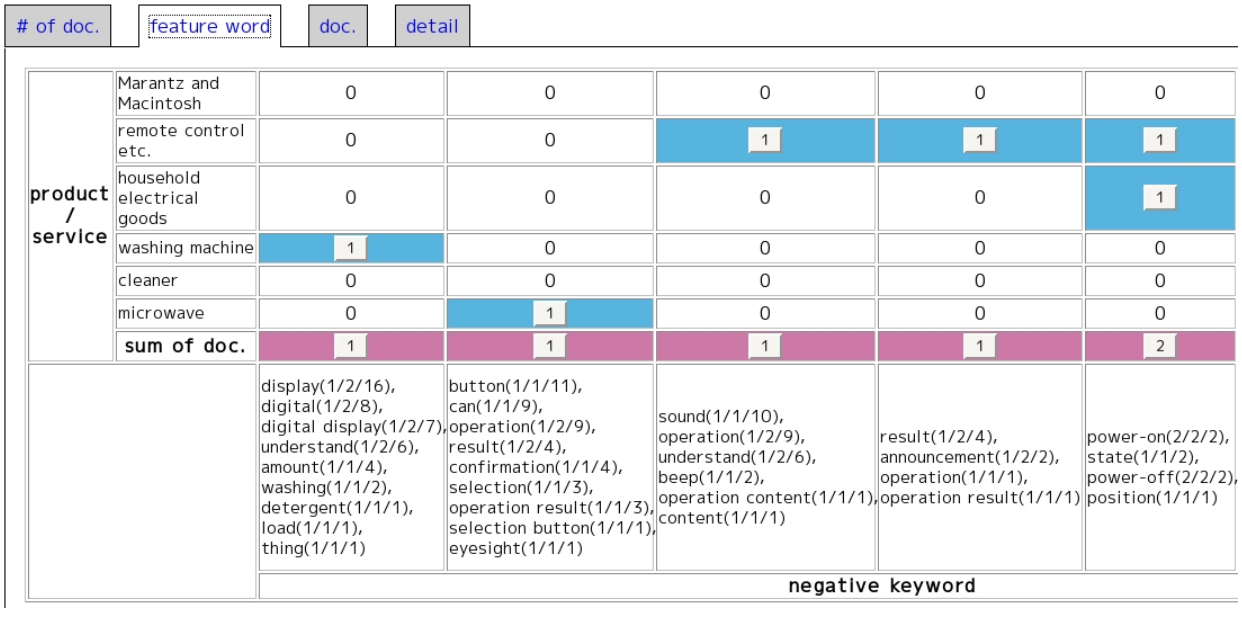

Fig. 3 The result of the retrieval for “switch” of the matrix search engine

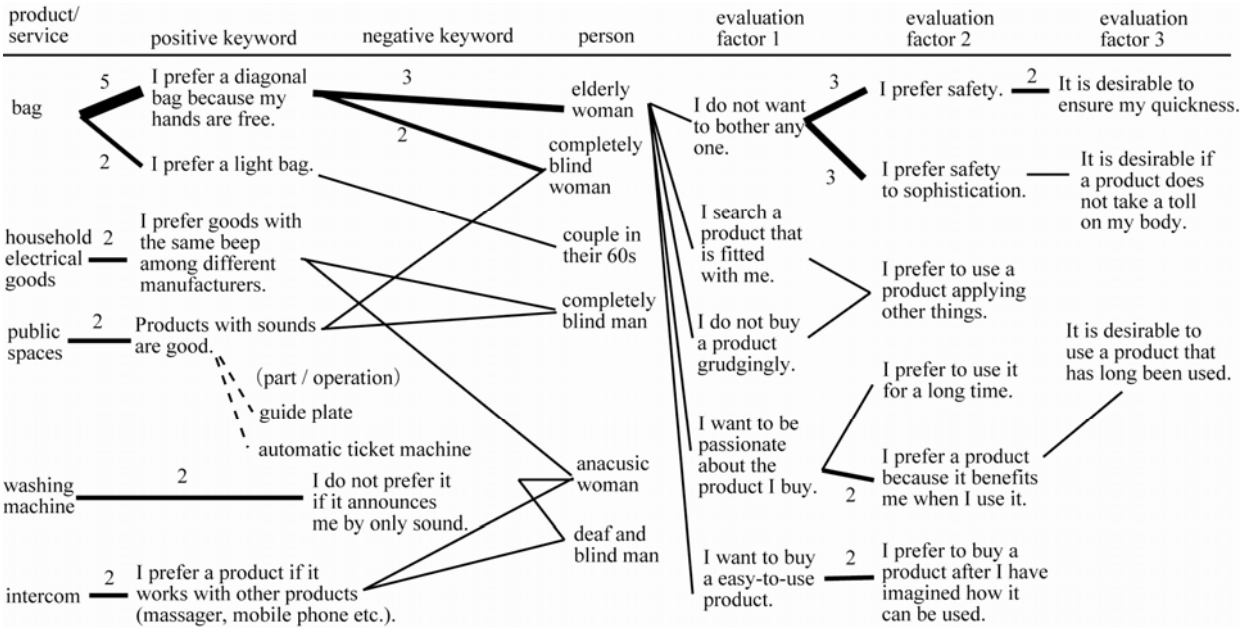

Fig.4 The example figure of analysis of evaluation records

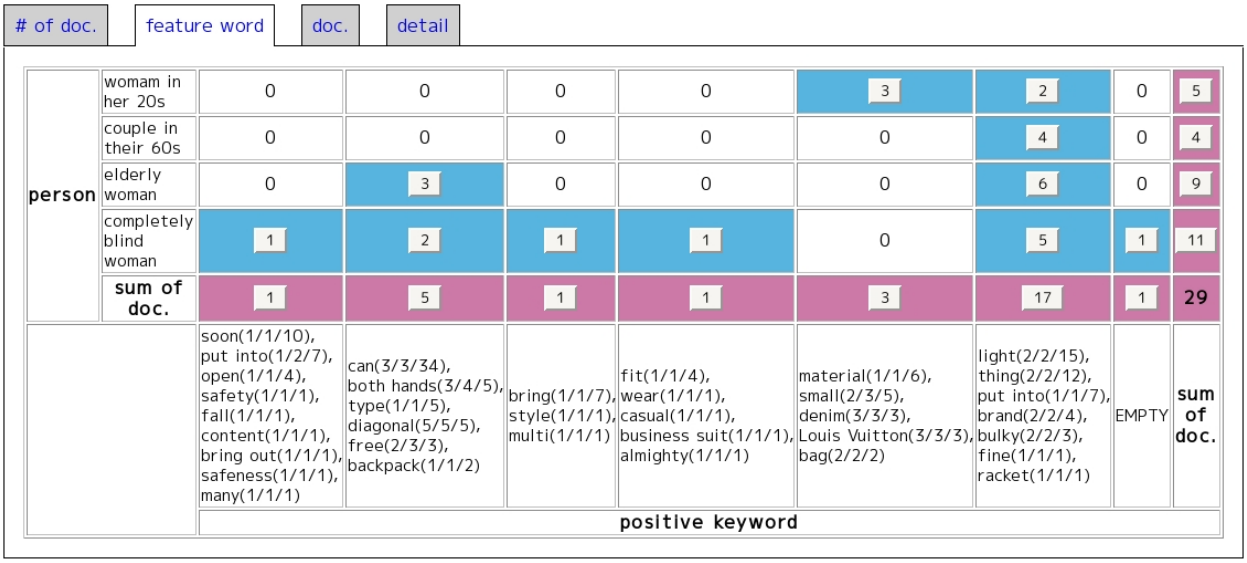

Fig. 5 The result of the retrieval for "bag"

the feature word is the document frequency of the word. The left number is the document frequency of the word in the cluster. The center number is the frequency in the retrieval records. The right number is the document frequency in all records.

The system can re-cluster retrieval records using two other aspects. For example, it re-clusters the retrieval records of Fig. 3 by using the two aspects, "person" and "evaluation factor 1." The system can also re-cluster only records which are assigned in a particular cell by clicking on the mouse when the cursor is in the cell. We can analyze retrieval records from various aspects in detail.

This system has been implemented for activities of academic staff [6], paper lists of The Institute of Electronics, Information and Communication Engineers [7], and ZOOLOGICAL SCIENCE, which is the official journal of the Zoological Society of Japan [8].

4. Experiments

This section shows the analysis of elderly and disabled peoples' evaluation records using the concept graph system and matrix search system. We made an analysis by combining results of the two systems (see Fig. 4).

4.1 Analysis of Product / Service

We found positive and negative keywords which are common to multiple persons when the target search aspect is "product and service."

We analyzed frequent products and services in all evaluation records by using a matrix search engine. The

Fig. 6 Part of the concept graph of 56 evaluation records in which the word "elderly woman" appears

search words were "mobile phone," "bag (29)," "wrist watch," "remote control," "button," "cleaner," "fax," "household electrical goods (9)," "shoes," "washing machine (8)," "living space," "announcement lamp," "toilet," "personal computer," "clothing," "TV," "intercom (5)" and "public space (5)." The underlined words meet the above condition. The number in brackets is the number of retrieval records.

Fig. 5 shows an output of the system when the search keyword is "bag," a row aspect is "person," and a column aspect is "positive keyword." The number of clusters of the row is 4 and the column number is 6 except for the "EMPTY" cluster. We see that the second cluster of "positive keyword" has the same need between elderly and woman with completely blind. The need is that they prefer a diagonal bag because both hands are free. By re-clustering the 17 records in the second cluster on the right of "positive keyword," we extract the need that elderly and women in their 60s prefer a light bag.

We made the left part, which is from "product / service" to "person," of Fig. 4 by combining the results of the above five words manually. The number on the line in Fig. 4 is the number of the records which are related to two aspects. A line without a number stands for one record. For instance, when the target is a bag, there exist five records which describe that some people prefer a diagonal bag because both of their hands are free and there exist two records which describe that they prefer a light bag. We can understand the amount of need because the cell displays the number of records. In the abovementioned records, the number of records of an elderly woman is three and that of a woman with completely blind is two. Thus, we can determine the relation of "product / service," "positive keyword," "negative keyword" and "person."

We can analyze retrieval records from different views by changing the aspects of a row and a column. We see that there exist two records for public spaces in Fig. 4. We changed the clustering target aspect "person" to "part / operation." We discovered that the targets of the records are a guide plate and an automatic ticket machine in public spaces.

4.2 Analysis of Person

After the analysis of Sec. 4.1, we analyzed 56 evaluation records of an elderly woman by using evaluation factors 1, 2 and 3.

Fig. 6 shows part of the concept graph of the records. We see that "attachment" and "shoes" appear as a lower concept of "elderly woman." We also see that the word

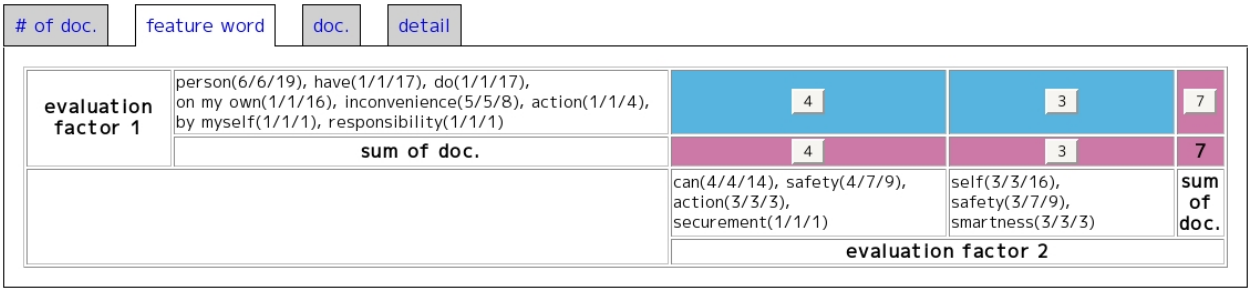

Fig. 7 The result of retrieval for "elderly woman"

"diagonal" is a lower concept of "elderly woman." This feature appears as a positive keyword of bag in Fig. 4.

Next, we use the matrix search system to analyze the records. The search keyword is "elderly woman" and the target search area is "person." We selected two aspects from evaluation factors 1, 2 and 3. The right part of Fig.4 was generated by combining the results of the system. Fig. 7 shows the result obtained by changing the number of clusters and selecting a particular cell. We see that the words "by myself," "action" and "inconvenience" in the aspect "evaluation factor 1" is related to "safety" and "smartness" in "evaluation factor 2."

We analyzed the evaluation records for two different problems in Sec. 4.1 and Sec. 4.2. We made Fig. 4 by combining the results of these problems. The two systems achieved dialogical analysis of the records. This enables us to examine various problems that we choose to target.

A problem of the concept graph system is that the generated graph is too simple if the evaluation record is short or the number of retrieval records is small. The generated graph often has only one-level depth. Such a graph shows only the co-occurrence of words. Consequently, we cannot determine the relation of feature words in detail.

A problem of the matrix search engine is that the numbers of records of clusters vary greatly. For example, the number of records of some cluster may be quite large but that of other clusters may be quite small. We cannot know the relation between two aspects in this situation. For example, there exist clusters whose number of records is 1 in Fig. 5. Such a record is independent from other records. In this case, the record is assigned to an "other" cluster. An important future work is to develop a clustering method to assign the averages of records.

5. Conclusion

This paper discussed a method to analyze the remarks of elderly and disabled people by using two text mining systems: the concept graph system and matrix search system. The two systems achieved multiple and dialogical analysis of the remarks. Also, the systems enabled us to examine various problems that we targeted. It is clear that the systems can apply to other documents.

We transformed the remarks to evaluation records with ten aspects to use with the matrix search system. It is necessary to improve the system to analyze original remarks because it is hard work to transform the remarks manually. We think the system can analyze original remarks if we have a list of words for each aspect. The words are used as word vectors for clustering.

6. Acknowledgment

We would like to thank the experimental subjects and the Accessible Design Foundation of Japan for their cooperation in this research.

This research is supported by the Japan Science and Technology Agency's project to develop "Innovative Seeds" in Fiscal 2006 and Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

7. References

1. J. Sanui and M. Inui. PHENOMENOLOGICAL APPROACH TO THE EVALUATION OF PLACES : A study on the construct system associated with place evaluation : 1. Journal of architecture, planning and environmental engineering, Transactions of AIJ, 367, 15-22 (1986). (in Japanese)

2. Y. Ohsawa, N. E. Benson and M. Yachida. KeyGraph: Automatic Indexing by Co-Occurrence Graph Based on Building Construction Metaphor. Proc. of the 5th International Forum on Research and Technology

Advances in Digital Libraries, 12-18 (1998).
3. T. Aikawa, T. Itoh, Y. Takayama, K. Suzuki and M. Imamura. A Proposal of a Method of Analysis of Questionnaires Using Text Mining Based on Concept Extraction. IPSJ SIG Notes, 2003-FI-70-1, 1-6 (2003). (in Japanese)
4. S. Hirokawa, E. Itoh, Y. Shimoji, Y. Dou and D. Ikeda. Relation Extraction of Songs and Artists from iPod Playlists. IPSJ SIG Notes, 2005-FI-81-(4), 31-36 (2005). (in Japanese)
5. S. Hirokawa, Y. Shimoji and T. Wada. Construction of Concept Graph from Documents. IPSJ SIG Notes, 2005(49), 79-84 (2005). (in Japanese)
6. S. Hirokawa, T. Seki, Y. Yasumoto and Y. Yamada. Faceted Search Engine for Activities of Academic Staff. IPSJ SIG Notes, 2005-DBS-137(2)-(89), 665-672 (2005). (in Japanese)
7. T. Seki, T. Wada, Y. Yamada and S. Hirokawa. A Faceted Search Engine for Supporting Information Retrieval and Analysis. The 18th IEICE Data Engineering Workshop, 2007. (in Japanese)
8. T. Seki, T. Wada, Y. Yamada, N. Ytow and S. Hirokawa. Multiple Viewed Search Engine for e-Journal - a Case Study on Zoological Science. The 12th International Conference on Human-Computer Interaction, 2007.
9. S. Chaudhuri and U. Dayal. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record, 26(1), 65-74 (1997).
10. K. Kato, Y. Yamada, S. Hirokawa and Y. Morita. User Assessment for the Formation of Assessment Indicators in Universal Design. Proc. (CD-ROM) of the 2nd International Conference for Universal Design, (2006).